

```

1 CAACCATTCG AGATCAGTGT GTGAGGGAAC TGCCATCATG AGGTCTGACA
51 AGTCAGCTTT GGTATTTCTG CTCTGCAGC TCTTCTGTGT TGGCTGTGGA
101 TTCTGTGGGA AAGTCCTGGT GTGGCCCTGT GACATGAGCC ATTGGCTTAA
151 TGTCAAGGTC ATTCTAGAAG AGCTCATAGT GAGAGGCCAT GAGGTAACAG
201 TATTGACTCA CTCAAAGCCT TCGTTAATTG ACTACAGGAA GCCTTCTGCA
251 TTGAAATTTG AGGTGGTCCA TATGCCACAG GACAGAACAG AAGAAAATGA
301 AATATTTGTT GACCTAGCTC TGAATGTCTT GCCAGGCTTA TCAACCTGGC
351 AATCAGTTAT AAAATTAAAT GATTTTTTTG TTGAAATAAG AGGAACTTTA
401 AAAATGATGT GTGAGAGCTT TATCTACAAT CAGACGCTTA TGAAGAAGCT
451 ACAGGAAACC AACTACGATG TAATGCTTAT AGACCCTGTG ATTCCTGTG
501 GAGACCTGAT GGCTGAGTTG CTTGCAGTCC CTTTGTGTGT CACACTTAGA
551 ATTTCTGTAG GAGGCAATAT GGAGCGAAGC TGTGGGAAAC TTCCAGCTCC
601 ACTTTCTAT GTACCTGTGC CTATGACAGG ACTAACAGAC AGAATGACCT
651 TTCTGGAAG AGTAAAAAAT TCAATGCTTT CAGTTTTGTT CCACTTCTGG
701 ATTCAGGATT ACGACTATCA TTTTGGGAA GAGTTTTATA GTAAGGCATT
751 AGGAAGGCCCT ACTACATTAT GTGAGACTGT GGGAAAAGCT GAGATATGGC
801 TAATACGAAC ATATTGGGAT TTTGAATTTT CTCAACCATA CCAACCTAAC
851 TTTGAGTTTG TTGGAGGATT GCACTGTAAC CCTGCCAAAG CTTTGCCATA
901 GGAATGGAA AATTTTGTC AGAGTTCAGG GGAAGATGGT ATTGTGGTGT
951 TTTCTCTGGG GTCAGTGTTC CAAATGTTA CAGAAGAAAA GGCTAATATC
1001 ATTGCTTCAG CCCTTGCCCA GATCCACAG AAGGTGTTAT GGAGGTACAA
1051 AGGAAAAA CCATCCACAT TAGGAGCCAA TACTCGGCTG TATGATTGGA
1101 TACCCAGAA TGATCTTCTT GGTCAATCCA AAACCAAAGC TTTTATCACT
1151 CATGGTGGA TGAATGGGAT CTATGAAGCT ATTTACCATG GGGTCCCTAT
1201 GGTGGGAGTT CCCATATTTG GTGATCAGCT TGATAACATA GCTCACATGA
1251 AGGCCAAAGG AGCAGCTGTA GAAATAAACT TCAAAACTAT GACAAGCGAA
1301 GATTTACTGA GGGCTTTGAG AACAGTCATT ACCGATTCTT CTTATAAAGA
1351 GAATGTATG AGATTATCAA GAATTCACCA TGATCAACCT GTAAAGCCCC
1401 TAGATCGAGC ACTCTTCTGG ATCGAGTTTG TCATGCGCCA CAAAGGAGCC
1451 AAGCACCTGC GATCAGCTGC CCATGACCTC ACCTGGTTCC AGCACTACTC
1501 TATAGATGTG ATTGGGTTCC TGCTGACCTG TGTGGCAACT GCTATATTCT
1551 TGTTACAAA ATGTTTTTTA TTTTCCTGTC AAAAATTTAA TAAACTAGA
1601 AAGATAGAAA AGAGGGAATA GATCTTTCCA AATCAAGAA AGACCTGATG
1651 GGGTAATCCT GTTAATTTCA GCCACATAGA ATTTGGTGAA AACCTTGCTA
1701 TTTTCATATT ATCTATTCTG TTATTTTATC TTAGCTATAT AGCCTAGAAT
1751 TCCATGATCA TGAGGTTGTG AGTATATCTC ATTTCTTCGT TGCATTTTCC
1801 TAGGTGTGCT TACTCTCTTC TCTCACTTTG TGACACAAGG ACATGAATAC
1851 ATCTAAATTT TCCTATTTCT GATATCACTG TTCCATGAC GTCATTAATT
1901 CTCTAACCTT AAGTGATAGG GTGACCTGCA ATATGCTGAT TCCTGGTGT
1951 TGCACAAACA CTGGATGTA AAGAAGTAAA AAATGTAAAA TTCACAAAAT
2001 TCAGTAAACC ACACAAATCA ATGAAGCATT CTATGACATT AGCTTGTTAT
2051 GAGTAACATA ATGATTTTTT TTTTCAATT TAAATAAGCC CTTCTACATA
2101 CCCAGCATT CTGATCTCAG ACAATGAATT GCTAAAAATG ACGATAGGGC
2151 ATTACACTCA GAATAGTTTG CTATATTTCC ACATACCTCA TCTAGATGTC
2201 ATAGCCTACA TTTCTGCCAT CACTTAACTG ACATTTTTTG TGTGTTCTTG
2251 ATGATAAATA GACAGTTCTT ATTATTGTCC TCAAATAATA AAAGAACTG
2301 AAATTTTCTT ACATAGAGAA AATGTCCATA AGATATTCAA GTTAAACAGA
2351 TTATTTTGAG ATAAGTAACC ATTAGAAATA TGTGATTGTA ATTTCTGATT
2401 TTATAAAATT TTAATTGATA GTACACTTGA TTTAAATGTC TATTCTTTAA
2451 AATGATGAAT ACTCATAATT CTTATCTCTA TAATCAAAAG TATAATTTAC
2501 TGTAGAAAAA TAAAGAGATG CTTGTTCTGA AAGTAAAAAA AAAAAAAAAA
2551 AAAAAGCTGT CATGCCGTTA CGTAGCGTAT CGTTGACAGC CCACTGTCAT
2601 GCCGTTACGT AGCATATCGT TGACAGCGAC ACTGTCATGC CGTTACGTAG
2651 CGTATCGTTG ACAGCACTGT CATGCGTTAC GAGCGTATCG TTGACAGCAC
2701 TGTCATGCCG TTACGTAGCG TATCGTTGAC AGCAAAACAC TGTCAGCCGT
2751 TACGTAGCG (SEQ ID NO:1)

```

FEATURES:

5'UTR: 1-37
Start Codon: 38
Stop Codon: 1619
3'UTR: 1622

Homologous proteins:

Top 10 BLAST Hits

	Score	E
CRA 147000022596013 /altid=gi 10438148 /def=dbj BAB15179.1 (AK...	931	0.0
CRA 1000682322899 /altid=gi 5802604 /def=gb AAD51732.1 (AF1752...	795	0.0
CRA 335001098690982 /altid=gi 11436851 /def=ref XP_003547.1 UD...	679	0.0
CRA 18000005226060 /altid=gi 5803213 /def=ref NP_006789.1 UDP ...	677	0.0
CRA 18000005155438 /altid=gi 4507821 /def=ref NP_001068.1 UDP ...	676	0.0
CRA 18000005147363 /altid=gi 6175083 /def=sp P06133 UDB4_HUMAN ...	675	0.0
CRA 18000004953169 /altid=gi 484383 /def=pir JN0619 glucuronos...	674	0.0
CRA 18000005148770 /altid=gi 3153832 /def=gb AAC95002.1 (AF064...	674	0.0
CRA 18000004903910 /altid=gi 4507819 /def=ref NP_001067.1 UDP ...	669	0.0
CRA 18000005164461 /altid=gi 3426332 /def=gb AAC32272.1 (AF081...	668	0.0
CRA 1000682327923 /altid=gi 5881246 /def=gb AAD55093.1 AF180322...	668	0.0
CRA 18000005219476 /altid=gi 8134780 /def=sp Q9XT55 UDBJ_MACFA ...	667	0.0

BLAST dbEST hits:

gi 10971169 /dataset=dbest /taxon=96...	383	e-103
gi 11068678 /dataset=dbest /taxon=96...	234	6e-59
gi 679005 /dataset=dbest /taxon=9606 /...	212	2e-52
gi 3173232 /dataset=dbest /taxon=9606 ...	212	2e-52
gi 3134358 /dataset=dbest /taxon=9606 ...	212	2e-52
gi 10298020 /dataset=dbest /taxon=96...	200	8e-49
gi 11974507 /dataset=dbest /taxon=96...	196	1e-47
gi 11973717 /dataset=dbest /taxon=96...	172	2e-40
gi 12673874 /dataset=dbest /taxon=96...	137	1e-29
gi 10887798 /dataset=dbest /taxon=96...	125	4e-26

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

Expression information from BLAST dbEST hits:

gi|10971169 Kidney-hypernephroma
gi|11068678 HepG2 cell line
gi|679005 Liver
gi|3173232 Kidney
gi|3134358 Kidney
gi|10298020 Hepatocellular carcinoma
gi|11974507 Normal pigmental retinal epithelium
gi|11973717 Normal pigmental retinal epithelium
gi|12673874 Kidney hypernephroma
gi|10887798 Kidney

Expression information from PCR-based tissue screening panels:

Human fetal liver

1 MRSDKSALVF LLLQLFCVGC GFCGKVLVWP CDMSHWLNVK VILEELIVRG
 51 HEVTVLTHSK PSLIDYRKPS ALKFEVVHMP QDRTEENEIF VDLALNVLP
 101 LSTWQSVIKL NDFEVEIRGT LKMMCESFIY NQTLMKKLQE TNYDVMLIDP
 151 VIPCGDLMAE LLAVPFVLTLL RISVGGNMER SCGKLPAPLS YVPVPMGTGLT
 201 DRMTFLERVK NSMLSVLEHF WIQDYDYHFW EEFYSKALGR PTTLCETVGK
 251 AEIWLIRTYW DFEFPQPYQP NFEFVGGLHC KPAKALPKEM ENFVQSSGED
 301 GIVVFSLGSL FQNVTEKAN IASALAQIP QKVLWRYK GK KPSTLGANTR
 351 LYDWIPQNDL LGHPKTKAFI THGGMNGIYE AIYHGVPMVG VPIFGDQLDN
 401 IAHMKAKGAA VEINFKTMTS EDLLRALRTV ITDSSYKENA MRLSRIHHDQ
 451 PVKPLDRAVF WIEFVMRHKG AKHLRSAHD LTWFQHYSID VIGFLLTCVA
 501 TAIFLFTKCF LFSCQKFNKT RKIEKRE (SEQ ID NO:2)

FEATURES:

Functional domains and key regions:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
 N-glycosylation site

Number of matches: 3

1	131-134	NQTL
2	313-316	NVTE
3	518-521	NKTR

[2] PDOC00004 PS00004 CAMP_PHOSPHO_SITE
 CAMP- and cGMP-dependent protein kinase phosphorylation site

Number of matches: 2

1	67-70	RKPS
2	340-343	KKPS

[3] PDOC00005 PS00005 PKC_PHOSPHO_SITE
 Protein kinase C phosphorylation site

Number of matches: 6

1	3-5	SDK
2	120-122	TLK
3	169-171	TLR
4	200-202	TDR
5	435-437	SYK
6	520-522	TRK

[4] PDOC00006 PS00006 CK2_PHOSPHO_SITE
 Casein kinase II phosphorylation site

Number of matches: 9

1	62-65	SLID
2	141-144	TNYD
3	204-207	TFLE
4	243-246	TLCE
5	258-261	TYWD
6	296-299	SSGE
7	297-300	SGED
8	419-422	TSED
9	435-438	SYKE

[5] PDOC00007 PS00007 TYR_PHOSPHO_SITE
 Tyrosine kinase phosphorylation site

Number of matches: 2

1	122-130	KMMCESFIY
2	136-143	KKLQETNY

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[6] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 4
1 19-24 GCGFCG
2 276-281 GGLHCK
3 373-378 GGMNGI
4 377-382 GIYEAI

[7] PDOC00009 PS00009 AMIDATION
Amidation site
338-341 KGKK

[8] PDOC00359 PS00375 UDPGT
UDP-glycosyltransferases signature
354-397 WIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPVMGVPIFGDQ

[9] PDOC00804 PS01047 HMA
Heavy-metal-associated domain
12-41 LLQLFCVCGFCGKVLVWPCDMSHWLNVKV

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	5	25	1.802	Certain
2	157	177	0.765	Putative
3	181	201	0.779	Putative
4	377	397	0.735	Putative
5	491	511	1.931	Certain

BLAST Alignment to Top Hit:

```
>CRA|147000022596013 /altid=gi|10438148 /def=dbj|BAB15179.1|
      (AK025587) unnamed protein product [Homo sapiens]
      /org=Homo sapiens /taxon=9606 /dataset=nraa /length=449
      Length = 449
```

Score = 931 bits (2381), Expect = 0.0
Identities = 448/449 (99%), Positives = 448/449 (99%)

```
Query: 79  MPQDRTEENEIFVDLALNVLPGLSTWQSVIKLNDFFVEIRGTLKMMCESFIYNQTLMKKL 138
      MPQDRTEENEIFVDLALNVLPGLSTWQSVIKLNDFFVEIRGTLKMMCESFIYNQTLMKKL
Sbjct: 1  MPQDRTEENEIFVDLALNVLPGLSTWQSVIKLNDFFVEIRGTLKMMCESFIYNQTLMKKL 60

Query: 139 QETNYDVMLIDPVI PCGDLMAELLAVPFVLT LRISVGGNMERS CGKLPAPLSYVPVPM TG 198
      QETNYDVMLIDPVI PCGDLMAELLAVPFVLT LRISVGGNMERS CGKLPAPLSYVPVPM TG
Sbjct: 61  QETNYDVMLIDPVI PCGDLMAELLAVPFVLT LRISVGGNMERS CGKLPAPLSYVPVPM TG 120

Query: 199 LTDRMTFLERVKNSMLS SVLFHFHWIQDYDYHFWEEFYSKALGRPTTLCETVGKAEIWLIRT 258
      LTDRMTFLERVKNSMLS SVLFHFHWIQDYDYHFWEEFYSKALGRPTTLCETVGKAEIWLIRT
Sbjct: 121 LTDRMTFLERVKNSMLS SVLFHFHWIQDYDYHFWEEFYSKALGRPTTLCETVGKAEIWLIRT 180

Query: 259 YWDFEFPPQPYQPNFEFVGGLHCKPAKALPKEMENFVQSSGEDGIVVFSLGS L FQNVTEEK 318
      YWDFEFPPQPYQPNFEFVGGLHCKPAKALPKEMENFVQSSGEDGIVVFSLGS L FQNVTEEK
Sbjct: 181 YWDFEFPPQPYQPNFEFVGGLHCKPAKALPKEMENFVQSSGEDGIVVFSLGS L FQNVTEEK 240

Query: 319 ANIIASALAQIPQKVLWRYKGKKPSTLGANTRLYDWIPQNDLLGHPKTKAFITHGGMNGI 378
      ANIIASALAQIPQKVLWRYKGKKPSTLGANTRLYDWIPQNDLLGHPKTKAFITHGGMNGI
Sbjct: 241 ANIIASALAQIPQKVLWRYKGKKPSTLGANTRLYDWIPQNDLLGHPKTKAFITHGGMNGI 300

Query: 379 YEAIYHGVP MVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLRALRTVITDSSYKE 438
      YEAIYHGVP MVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLRALRTVITDSSYKE
Sbjct: 301 YEAIYHGVP MVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLRALRTVITDSSYKE 360

Query: 439 NAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTW FQHYSIDVIGFLLTC 498
      NAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTW FQHYSIDVIGFLL C
Sbjct: 361 NAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTW FQHYSIDVIGFLLAC 420

Query: 499 VATAIFLFTKCFLESCQKF NKTRKIEKRE 527 (residues 79-527 of SEQ ID NO:2)
      VATAIFLFTKCFLESCQKF NKTRKIEKRE
Sbjct: 421 VATAIFLFTKCFLESCQKF NKTRKIEKRE 449 (SEQ ID NO:4)
```

```
>CRA|1000682322899 /altid=gi|5802604 /def=gb|AAD51732.1| (AF175221)
      UDP glucuronosyltransferase UGT2A3 [Cavia porcellus]
      /org=Cavia porcellus /taxon=10141 /dataset=nraa
      /length=530
      Length = 530
```

Score = 795 bits (2030), Expect = 0.0
Identities = 377/530 (71%), Positives = 435/530 (81%), Gaps = 3/530 (0%)

```
Query: 1  MRSDKSALVFLLLQLFCVCGFCGKVLVWPCDMSHWLNKVKVILEELIVRGHEVTVLTHSK 60
      M K A LLL L C G GFCGKVLVWPC+MSHWLN+K +LEEL+ RGHEVTVL T S
Sbjct: 1  MAPGKLASAVLLLLLCCAGSGFCGKVLVWPCEMSHWLNKLTLLLEELVKRGHEVTVL T LSN 60

Query: 61 PSLIDYRKPSALKFEVVHMPQDRTEENEI---FVDLALNVLPGLSTWQSVIKLNDFFVEI 117
      IDY + A FEV+ +P D+ I F++LA+NV+P + WQS L FFV+I
Sbjct: 61 NLFIDYNRHPAFNFEVIVPPTDKNMSENILNEFIELAVNVMPMTPLWQSGKLLQQFFVQI 120

Query: 118 RGT LKMMCESFIYNQTLMKKLQETNYDVMLIDPVI PCGDLMAELLAVPFVLT LRISVGGN 177
      L + C + +YNQ+LMKKL+++ YDV++ DPVIPCGL+AE+L VPFV L+ S+G
Sbjct: 121 TEDLGLNCRNTVYNQSLMKKL RDSKYDVLVTD PVI PCGELVAEMLGVFFVNMLKFSMGHT 180
```

FIGURE 2, page 3 of 4

Query: 178 MERSCGKLPAPLSYVPVPMGLTDRMTFLERVKNSMLSVLFHFWIQDYDYHFWEEFYKA 237
+E+ CG+LPAP SYVPVP+ GLT RMTF+ERVKN + SVLF FWIQ YDY FW++FYS+A
Sbjct: 181 IEKYCGQLPAPPSYVPVPLGGLTTRMTFMERVKNMVFSVLFDFWIIQQYDYKFWDQFYSEA 240

Query: 238 LGRPTTLCETVGKAEIWLIRTYWDFEFPQPYQPNFEFVGGLHCKPAKALPKEMENFVQSS 297
LGRPTTLCE +GKAEIWLIRTYWDFEFP+PY PNFEFVGGLHCKPAK LPKEME FVQSS
Sbjct: 241 LGRPTTLCEIMGKAEIWLIRTYWDFEFPRLPNFEFVGGLHCKPAKPLPKEMEEFVQSS 300

Query: 298 GEDGIVVFSLSLGFQNVTEEKANIIASALAQIPQKVLWRYKGKKPSTLGANTRLYDWIPQ 357
GEDG+VVFSLGS+ +N+TEEKAN+IASALAQIPQKVLWRYKGKKP+TLG NTRL+DWIPQ
Sbjct: 301 GEDGVVVFSLGSMVKNLITEKANLIASALAQIPQKVLWRYKGKKPATLGPNTRLFDWIPQ 360

Query: 358 NDLLGHPKTKAFITHGGMNGIYEAIYHGVPVMVGPIFGDQLDNIAHMKAKGAHAVEINFKT 417
NDLLGHPKTKAFITHGG NGIYEAIYHGVPVMVG+PIF DQ DN+A MKAKGAHAVE+N T
Sbjct: 361 NDLLGHPKTKAFITHGGSNGIYEAIYHGVPVMVGMPIFSDQPDNLAGMKAKGAHAVEVNMNT 420

Query: 418 MTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSA 477
MTS DLL ALRTVI D +YKENAM+LSRIHHDQPVKPLDRA FW+EFVM HKGAKHLR A
Sbjct: 421 MTSADLLGALRTVINDPTYKENAMKLSRIHHDQPVKPLDRAAFWVEFVMHHKGAKHLRVA 480

Query: 478 AHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE 527 (residues 1-
527 of SEQ ID NO:2)
AHDL+WFQ++S+DVIGFLL CVA+AI L TKC LFS Q F K K K+E
Sbjct: 481 AHDLSWFQYHSLDVIGFLLACVASAILLVTKCCLFSFQNFQIKIGKRIKKE 530 (SEQ ID NO:5)

Model	Description	Score	E-value	N
PF00201	UDP-glucuronosyl and UDP-glucosyl transferas	962.0	1.5e-285	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF00201	1/1	24	525 ..	1	507 []	962.0	1.5e-285

```

1 TTCTAGAGGG TTGAACAAC TTTCCCTGA TACATTGCAT TTTTGTGATA
51 CCTTCAGTAC ATGTTAAACT GGCAACCACC AGTGAACCTT ACTCTTAAAA
101 TATTAATTTT TAACCTCTGT GCTTATATTG TCATTTCAAC TCCTTGCTTA
151 GTAACACAA AACCATTGCA GATCAGTGTG TGAGGGAACCT GCCATCATGA
201 GGTCTGACAA GTCAGCTTTG GTATTTCTGC TCCTGCAGCT CTTCTGTGTT
251 GGCTGTGGAT TCTGTGGGAA AGTCCTGGTG TGGCCCTGTG ACATGAGCCA
301 TTGGCTTAAT GTCAAGGTCA TTCTAGAAGA GCTCATAGTG AGAGGCCATG
351 AGGTAACAGT ATTGACTCAC TCAAAGCCTT CGTTAATTGA CTACAGGAAG
401 CCTTCTGCAT TGAAATTTGA GGTGGTCCAT ATGCCACAGG ACAGAACAGA
451 AGAAAATGAA ATATTTGTTG ACCTAGCTCT GAATGTCTTG CCAGGCTTAT
501 CAACCTGGCA ATCAGTTATA AAATTAATG ATTTTTTTGT TGAAATAAGA
551 GGAACTTTAA AAATGATGTG TGAGAGCTTT ATCTACAATC AGACACTTAT
601 GAAGAAGCTA CAGGAAACCA ACTACGATGT AACGCTTATA GACCCTGTGA
651 TTCCCGTGG AGACCTGATG GCTGAGTTGC TTCCAGTCCC TTTTGTGCTC
701 ACACCTAGAA CTTCTCTAAG AGGCAATATG GAGCGAAGCT GTGGGAAACT
751 TCCAGCTCCA CTTTCCTATG TACCTGTGCC TATGACAGGA CTAACAGACA
801 GAATGACCTT TCGGGAAGA GTAAAAAATT CAATGCTTTT AGTTTTGTTC
851 CACTTCTGGA TTCAGGATTA CGACTATCAT TTTTGGGAAG AGTTTTATAG
901 TAAGGCATTA GGTAAGACAC TTTTGTTTTA TTTTAAATTT AGTTATCAAA
951 AGAAATATTT TTAATAATTG TCATACATTG TCTATGACAT ATATATGCAG
1001 GTCATGAGT TTTTTTAGAA AATGTTGTAG CTGTTTTTCA TAAAGAAAGT
1051 GTATTTGTTT TAAGCGTAAG ATAACCTACT TTCTTAATAC CAGTAATATA
1101 CTTAAAAATG ATCATCAATA ACTAAGAGAT TATATTTTGT ATTTCCCTCA
1151 AATAGCGCAA ATCAACATCA CATATTTTGT AGAATCACTG ATTGTTAGTC
1201 TGAATGTTAT AGAATTTCTA TTGAAATAAA ATGCTAATCA TTATTTCTC
1251 TCTCATCATG TATTTAAGAA AATCTTCAGA AGGTCTTCTT TGAATTAATT
1301 TTTCAAGAGT CATTAATTTG AACATTTTCT AGAATTTCTT AATTTCTTAG
1351 GTGATTACTT CACAAAAACT TGAAAAATA TTATAAAAAG TTAAAAAACT
1401 TACGGTCTTG TGGGCGATAA GATAGTAGAA TTTTACTTTT ACTGATATAC
1451 ACCTATTTGA CTTATTTTTA TTTCTTTGCT TTACTGATAA AAAGTTGTTT
1501 TGCTTTGCAA TTTTCATATA GTTGTGATCA GAGCTGGTCA ATGCAAGACA
1551 TGTTTTTATC CAAATATGTT TGAGAAATAT GTAGAAACAT GAAAAAAGGT
1601 ACAATTATAT CCGACACTAA AATATTGTTT AATGTATTCC AACGAATFCT
1651 TATGCATAGA CTGTTTCACA GAACTAATAT TCAGAGGATC CCAGTTCAAA
1701 TGTCCCTTAGC CTTAGACATG ATTTGAATTT ACATATATTG ATTTGCTTTA
1751 AATAATTTTC CATTCAGTAA GCTGTGCCTA GCTGCAGATA GCCTACCAGG
1801 CTTTATGGAT CTAGGTAAAC AATACAAATC TCTTGGCCTC AAGTCTACAT
1851 TCAGATATTA ATTTAAAGGG GTACAGCTAT ATAGAGGTCA CTGGCAAATT
1901 TTGGTAAAT AGGATTATAG TAAAAGCCCC CTGACAAGAT TGAAATTTAA
1951 AATAAACAA AAGTGTATC AAAGGGGTGA AAGAGCATTT TCCAATAAAC
2001 AAAAGTGGG TCTGGCCATG CATTCAGAAA TTCCCAACA ATTCTTTAAA
2051 AATCATGGAG CAGCTTGATA TATAAGAAAT TCATTTAATA ACTATATTTA
2101 TTATGTAGCT CCAACTTACT AAATTATTGA TTATTATATA TTTTATAGAA
2151 TTATCTATTG TGAGTCTAAA TCAAGAGTAT ATATTCAAAC AACTATAGGA
2201 AAAGGGATAT CAGTCAATTT CAATTCAAGG ATTTATTTCC ATAAGTGCTT
2251 ACGCACAGGT GTATTTTATT TTATTATACA TTGCTTTATT GTCCTTACA
2301 AAAATTGCAA TTTACAAATT AAAGGTTTTT GAAAACCTTG AATCAAGCTA
2351 ATCAATTTGG CGTAATATTT CCAACAACAA GTGTGTACTT TTGACTCTAT
2401 CACATATTGG CATTTATCAT GCTTTTTCAA ATTTTTCATT GTTATATCTG
2451 TTACGGTGAT CTGGGATCAG GTTTCCTTGA TGGTACACG TTTATTAGCT
2501 TGGGGGCACC TTGATGTGTT ACAATATAAG ACAGCAAACCT TAATTATAAA
2551 TGTTGTGCAT GTACTAACTG CTCCGCTGAT TCGTTTCCCC ATCCCCTTC
2601 TTCTTAGGCC TCCCTATTCC CTGAGACACA GTAATATAAC ATACAATGAC
2651 TTCTAAATGT TCCAGTGAAA AGAAAAGTAG CAGGTCTCTC AATTTAAACC
2701 AAAAATATAA AGGAATAAGT TTAATGAGTA CTATAGTTTA GATATGGTTT
2751 GCTTGACCTT ACAAATCCT GTGTTGAAAT TTGATCACCA ATATTGGAGG
2801 TGGGGCTTGA TGGGAAGTGT TAGGGTCATG AGGGTAGATT CCTATGAGT
2851 ACATTAAATG TCTCCCTGGG GAAATGGGTG AGTTCGTTCT CACTCTATTA
2901 GGTCCCAGGA GAGATAATTA TTAATAAGAG CCAGGAACAT CCACCTCTT
2951 TCTCTTGCAT ATCTCTCAT TCTGTATCCC TGCACTTGCT GGCTCCCAAC
3001 ATCTTCTTCA ATGAGTGGAG GAAACCAGAG GTCTTCACCA GACACAGATG
3051 TTGGTGCCAT GCCTCTTGTA TACCCTGAAG AATTGTGAGC CAAATAAAAA
3101 CCTTTTTCTT TTACAAATTA GACAGCTCA GTTATTCCTT TGTAGCAACA

```

FIGURE 3, page 1 of 11

```

3151 AAAAAAGCCT GGGACAGGCC AAAAAGTTCCT CCATTGCACC AAACAGTTAA
3201 ACAAGATGTG AGTGCAAAGG AAAAGTTTTT GGAGGAAATT AAAAGTGCTA
3251 CTCCAGTGTA CATACAAATG ATAAGAACAA ATAACCATTA TCAGTGCTGA
3301 TATGGAGAAA ATTTTAGTTG TCTGGAGAGA AAATCAAATT AGCTAGCCAG
3351 CTGCAGTGAT TCATATCTGT AATCCCAGTA ACTTGGGAGG CTCAGGTGGG
3401 AGAACGGCTT GAGCCCAGAA GTTTGAAGTC CAAGGCTGCA GTGAGCTATG
3451 ATTGCTCCAC TGCACCTCAA CCTAGGTGAT AGAGCAAAAC CACTACCAAA
3501 AAAAAAAAAA AAAAAAGAA GAAAAAGAAA AGAAAAAAAA TTAAACCAAC
3551 CACAACATCA CCTTAGGTTT TGGCATTAGC TAAAACTAA TACATAGTAA
3601 AGCGTTAACT ATTCAATTGC ATGAAACCTC AGAGAGGAGA GGAAGATGCA
3651 GAAAAAAGA CTGAAGCTAG TAGAGGTTGA CTAATGAGGT TTACAGGAAT
3701 AACTGCCTA CATGATGCAA AAGTTCAATG TGAAGCAATA GGAAGTCATG
3751 CAGAGACTT AGCTAATATA CTCAGTAAAT GTGGCTACAG TAAACAAATG
3801 ATTTTCAATG TAGACCTAAC AGCCTTCTGT TGGAAGAAGA TGCCATTTAA
3851 AACTTTCATA GCTAGAGAAG AGAAGTCAAT GCTTGTCTCT GAAGCTACAA
3901 AAAACAGGCT GAATCTCTTG TAGTGGCTAA TGCAGCTGAT GACAAAGGTA
3951 AAGCCAATGC CCATTTACTT TTTGTAATAA TTATAGAGGA CTCTTAATAA
4001 TTATGTTAAA TCTACTTTGC CTGTGTTATA TCAATGGAAC AACAAAGCCT
4051 GGATGATATC ACATTGGTAT ATGACATGGC TTATTGAATA TTTTAAGCAC
4101 ACTGTTGAGA CCTATTGCTC AAAAAAGAGG ATTCCTTTCA AAATATTGCT
4151 GCTCATTGAC AATTACATG GTCAACAAAG GGCTCTGATT AAGATGTACA
4201 GATATTAATG TTTGCCTGCT TGCTATTATT ACATCCATCT TACATGCCAT
4251 GGATCATATA GCCTTGACTT TCAAGTCTTA TGTAAGAAAT ATATTTTGTA
4301 AGGCTATAGC TCTTACTAAT GGGGAAAGTA TATTGAAAAC CTTTTCAAAA
4351 GGATTTTCA TTCTAGATTC CATTAAGAAC ATTCATGGTT CATGAGAGGA
4401 AGTCAAGATA TTAACATTAA CAAGAGTTTG GAAAAAATTT GATTCTAACT
4451 CTCCTGGATG ATTTTGAGGG ATTGAAGACA TCATGTGAAG AATTAAGTGG
4501 GGATGGGGTG GTCATGAAAA AATAAATAGA ATTATAAGTG GGCCTGAAGG
4551 TTTGTCTAAA TTGCTATAAT ATCATGATAA AACTAAAACC TGTAAAACCG
4601 GTGAGGAGGT GCTTTTAAA CAGTTACTTT TTATAGATGA ACACAGAAAT
4651 TGGTTTTGTG AGTTGGAATC TTCTCCGAGT GAAAATGCTA TGAACATTGT
4701 TGAATGGCTT ACAAATGACT TAGAATATTA CACAAAATTA GTAGATAAGG
4751 CAGCATCAAG GTTTGAGAGA ATGGACTCAA ATTTTGAAAG AAATCTACT
4801 ATGGGTAAAC TGCTGTGAAA CATCATCATA TGCTACAGAG AAATCTTCA
4851 TGAAAAGATG AGTCAATTCA TGCAACAATC TTTGTTGTCT AATTTTAAAA
4901 ATTGTCCAGC TGCCCTGATC AATCAACAGT AATCAGCACT GAGGCAAGAC
4951 CCTACACCAG AAAAAATAA AATAAAAAAC CTCCTTGCT GAAGACTCAG
5001 CTTATTATTA GCACTTTTTA GCCATACTTT TAACTAAGGT ATGTGCATTC
5051 CTTTTTAAAC GTGATGATAT TGACACAGCTA ATAGCCTACA AGGTATGGTT
5101 AACATAACTT TTATATGTCC TGGGACCCAA ATTTGTGTGA ATCACTTTAT
5151 TGACATATTC CTTTATTGTA GATGAAGTGC AACTTATCTT GCAATATCTC
5201 CAAGATATGT TGTGATGGCA TTCAAATAA GATGTGAAAT TATTTTATTA
5251 GTATAAAAAG CAAATTTAAT TTTCTTCTCT TTGATCATCT TTATCCTTGT
5301 TACTGTGTAT TTATCCTTTA AACATTGAAT GACTCCAATT GTTTAAACT
5351 GAGTCTTTCT TAAATGAGTC CTAATATCAT AGTAATTAAA ATCACCTACA
5401 AGTTGGTAAT GCAGGACGCA TGTGAGGCAC AGAAAAACAAC AAATTTATAA
5451 GACATAAATG CATTTGCTTG GAAGCTGAGA GAAGGCTCTA TTCTAATTTT
5501 TGATAACTTC AAAGTGAGTA TCTTCAGTAA AATTTATTCA CTATCAAAAT
5551 CAAGCGCTTT GGATTTATGA CCTAGGAAAA AACTTCAAAC ATTAATAATG
5601 GATGACCTTA AAAAGAGGCT CTCCACACTA TGGTGTATAA CACCACCAAC
5651 TTTGATTAGA ATTTTAAAGA GAAACAAAT CTCTTATGGA GTTTATCTTT
5701 TTATCACTTG CAAAATATGT TTTTGTAAAG AGATACTAAT TACTTAGTTA
5751 TTTGTAGTTA GCCATTCTTC TGATTAATAA CCTAAAATTA AATCTTGAAA
5801 ATGTGTTTTT CTTCAAAACA CATCATTTGA GAGAAACACT AAAGTAAGTG
5851 TATGATTATC ATAGCATGTA CATAGGTGCT TCACAACCCA AAAAGAATAT
5901 TGTCAATGGT AAGAATCAGT AAAGGAATTT CTCCTAATAA AACAGTAGCC
5951 TATTAATTAA AGTAATGATA TGCAATACAG CAAGTTAAAG GGAAGTATC
6001 CTGGTGGGAT TATTGAAAGA TATACCTTTG ACTATAGATT AGAAAATACA
6051 GAGATGTTAT TTAGTGAAGA TATTGTGGTA CTCATTTATC ATCTGCAATT
6101 CACTTGCAGA GGAATAAATG AGTAATAAAT TCATTTGCAT TTTGGATTG
6151 TGTCTTTAAG TTGTGAAAA ACACCTTAAAT ATAACCATCT GTCCTTTGCT
6201 CCTTCCTTCC TTTCTCTTT CCTTCCTTCC TTTCTCTTT CCATCCTTCC
6251 CTCCCTCCAT CCTTCCTTCC TTTCTCTTT CCTTCCTTCC TTTCTTTCTC

```

FIGURE 3, page 2 of 11

6301 TGTCTTCCTT TCTTTTTTCC TTTCTTTCTT TTTCTTTCTT TATTATTTCA
6351 TTAATTTCCC CTTCATTG ACCTCTAAA GCCATGTTGT TCTAGAGGAC
6401 TTAAGCTTAT TTTTTTCTTA ATAGCTTACT GAAAAATTAG TGATACAATT
6451 TTTTATTTGA ATTGTATGCT AATTCATTCT GTTATTTCTT TTATTGAGGA
6501 AGGCCCCACTA CATTATGTGA GACTGTGGGA AAAGCTGAGA TATGGCTAAT
6551 ACGAACATAT TGGGATTTTG AATTTCTCTA ACCATACCAA CCTAACTTTG
6601 AGTTTGTGGG AGGATTGCAC TGTAAACCTG CCAAAGCTTT GCCTAAGGTA
6651 GGACTATTGT ATTAAGGAAT ATTATGTACT TTATGACATG ACTTGTTTTC
6701 CCTTGAAAGA TTACAACCTT AGTTATAGAA GGATGATGTT GAATGTCGTC
6751 TGTTTGACGC TCCATATTTA TTTTCCATGC CACAGGGGCT CTTATAGGTG
6801 ATTATATGTC TTTTCGGTAT TATATTGAGA AAGTAGGCAG AAGAATTTCA
6851 TGATTAGAAT AGATTTTAAA ATACTAGTAT TACAATAGTT TGGATAATAA
6901 ATTGAATTAA TAGGGAATTG GAGCCATGAA GATCACTAAA AAGAATGCTC
6951 TAGCCTTTCT CACAATCAAA TTGGGCTTAT GAACAAGGAT ATTTGTCTATG
7001 ATAGTACAGA AATAAGCATA TTTTCATGAG ACATATTGGA TATATTCCAC
7051 AGGAGTTGGT GAGTGAGAGA AAATAAGTGA TGAAGGAAGA CAAAGAATAA
7101 AAGAAAATTT CAATAAATGG AAAGTTAAG TGTTTAATGA TAGTGATGAC
7151 TTTTACTCAA ATAAGTGCTT AGAAGTCATC TTGTTTGTGA TTTATATGAT
7201 GAATTCTGTG TTGTGACTAT CCACTTTGAG CTCGTGAGAA TGTTAGGTGA
7251 GGTTTAATAA AAGCCATTG AGAAAAACAA GGTTTCAACC TCTGTGGACA
7301 GAAATCTAAA TATCGATAGT TATCAGGACA AAGTAGAGCT CATAGAAATA
7351 ATTTTGACGC CTGCAGGTTT GTTTTGGAGT GAAAAATAAA TTGTATACTA
7401 TATTCCTAAA TCATCAGAGG AAAAAATTTA TAGTTCAAGG AATGTTGAAA
7451 GAAACAATAT TGAGAAGTAA AAGTGAGTAA TAGTTGTTAT AGTTTTTTAA
7501 TAGTTTTGTA AGTATGTCTT GAGTTCCTG TCCCAAAAGT GGCTATTAGC
7551 TCTAGCCTTG ACCTGACAAG GTTCTAGGAT ATTTAGTCAT GGATGTTTCT
7601 AATCTACCTC TTACGGGATA CTTTTTATTC TGATGAACAG CCTAATGCCT
7651 AAGTGTGCAA TCTATACCAA GATTGTCTT ATAGGGAAC TGTTTACACT
7701 GGAAGACACC ACTGTGTCTC TTGTATGACC TATGTCTTCT TTATCCCTAC
7751 AAAGGTAACC ACATTATAGG AAACCCAGC AAGGCCAGAT GTTATATTTG
7801 TGTTGGTCAA GTGAGAAAAC ATGGGAGAAA CTTAACCAAA CACATAAAAT
7851 AACAGAAACA GTCTTCTTTG ACCATTTCTA GAGAAAAGAG TTCAGCATCC
7901 CTTGTAAGGC CACTAGGAAG AAGAAAATTC TCTGGGAAAA GCACATTCAA
7951 CCAATGAATG GAGACCAAGA AAGAGAGTGA GGGATCTATG TGCCAAAATG
8001 TTAAGTGGGA TCCAGGGTGT TACCTAGGTG GGTTTCCAAT GGGGAACGTG
8051 AATTGGTAGG TTTAATGCAA GCAGGCACAA AGTCCATGGA GGCATTCTGA
8101 GACTGAAAGA TAGTCACTTT GGCATATCTG CACAGAATCT GATCAGTGAT
8151 TCAAGCCCAA GTAGGCTGTA TCTAGTTGTC CTATAGGGTG GTTACCAGGA
8201 GGCAGTGTGT AAGTAAAAAT CCTGACTGAA CACATTGAGG AAATGGAAGG
8251 AGGTGGAAGA TTTTAAACGG TGTCAGTGTT GACTAAGACC TGCTTCTGGT
8301 ATGGAAAATT CAACTTATAT TTTAAATGCA TAGCCAGACA ACATAAAAT
8351 ATAAGAATTT ACCACAATAG CTATGGTAAC AATACTGGGT TTACCTATTA
8401 CTACAGAGTG AAAAGAAAAC CCTCATTTCC CATTTTATGG AAATATAATC
8451 AAAATCCTAT AAGGAAGGTT TCAGAGCCAG TAGGATTTCC AGAAAAATTA
8501 TTGGTTTTAT AGTAAGATGT GTATTGATGA ATATAATTTT ATTTATTAAT
8551 TATTAATATC ACTTTACTTA CCAGGAAAGT TATACCAGAA AACCAAGCTC
8601 TCTTAAGCCA TGGCATCTGT ATCTAAAATA GAAATACAGA AGGAGAGCTG
8651 ACAATTTCCA TCATTCTCTA GGTAATCTCC CATGCCATTC TACCCTTTAT
8701 TCCCACACTC CCAGTTTTAC ACACACACAC AAACACACAC ACACAAACAC
8751 ACACCTCATAG AAATAATCAT AGAAGACATA TTTTAAAAAA AGTTAGATCC
8801 ATACAGTAAT AATTTATTAG GTAAAAGCTT TTGTGCTGAT AATTTTACAA
8851 GTTTAATTGA GATATATTTT AGGGCTGTCT TACACTAAAT ATTTATTTT
8901 ATTTTAAAAA TTTGACATGT AATAATTGCA CATGTTTAAG AGAAATGCTG
8951 TGGTATTACA ATACATTTAA ATGTTGTGTA ATAATTACAT CAAGATAATA
9001 AACCACATCAT CTAATATTTT ATCATTCTCT TGTGGTGATA ACATTCAAAA
9051 ACCTCCTTTC TGGCTATCTT GAAATATGTA ATACATTACT ATTAACATA
9101 GTTACCCAAC AACTTAATAT AATAACAGAA CATATTCTTC CAAATTTAAA
9151 CGTTGTATCC ATTGATCCAC CATTTCTCAT TGCCCTCCCT ACTATCTCTT
9201 CAGCCTCTAG TAACCACAAT TCTACTCTCT AATTATATTA TGAATGCATT
9251 TTTTGATTCC ACATATAAGG GATACCATGC TATCTCTGCC TGGATTATTT
9301 CAGTTAACAT TATGCCCTGG AGGTTTCATT ATGTTTCTAG AAATGACAGG
9351 ATTTCAATTCT TTTTTCCTCA ATATATATTT AATGAAATGG ATATATATAA
9401 ACATTGGAAA ATGTATATAT ATATATATAT CTCCAGTGGA ATGCTATTGA

FIGURE 3, page 3 of 11

9451 GCTATAAAAA AGTTAATATA TAATAGAAAT AAAGCTTATA TATATCTAAT
 9501 GGAATGGATA TATATATATA ATGGAATAGA AATATATATC TATACATATA
 9551 AACACACGCA ATATACATAT CCATTTTCATT GCATATATAT ATATATAGAG
 9601 AGAGAGAGAG AGAGATATTT TCAAATGTGT GTATATATAT CCAATGGAAT
 9651 GGACATATAT ATATGTATAT TTTTCCATA TTTTCTTTAT GTATTTCTTC
 9701 ATTAATGGAT GTTTAGGTTG ATTCATCCCT TGGGTATATG AATAATGTTG
 9751 ATGTAAACAT AGAAGGACAG ATATCTCTAT GACTTCTTAG TTTATTTAAA
 9801 TATACACCCA GTAATGGAAA TGCTGTATAA TATGGTAGTT CTATTTTCAT
 9851 TTTTGGAGGA ACTACCATAC CGTTTTCCCTT ACTAATTGTA CTAATTTGCA
 9901 TTTCCCTCAA CAGTTTATAA AAGATCTTCT TTCTCTGCAT ACTTTCTAGC
 9951 ACTTGTTATT TTTGCCTTTT GATAATAGCC ATAACAGGGG TGATGTGATA
 10001 TCTCATTGTA GTTTTGATTT GCATTTCCCT GATGATTAGT GATTTTGAGC
 10051 ATTTTGTAAT TATACTTCTT AGTCACTGAT AGTCTTCTTT TGAGAAGTGT
 10101 CTATTCAGGT CTTTGTCTTA TTTTAAATC AAATTAGTAA TTTATTTTAA
 10151 TTGACTGATG TGACTTCTAT GTATATTTGA GATAGTAACT TATTGTCAGA
 10201 TTCATAGTTT GCAAATATTT TTCATGTTGT GAATTGTCTC TTCACCTGT
 10251 TGTTTGCTTC ATTTTCTCTG CACAAGCTCA ATGCTTTGAT ATAACCCATT
 10301 TATCTACTTT TCCTTTTGTG GGCTGTGCTT CTGAAGTCCT ATCCAAAAA
 10351 ATCCTTGCCT AGACCAATGT CACAAATCAT TCCTCCTACA GTTCTTCTA
 10401 GTAGTTGTAT AATGTTTGGC CTTATATTTA ACTTTGTAAT TCATTTTAC
 10451 TTACTTTGTA TATGGTGAGG GATAGAGGTC TAGTTTCATT TTCTGCATGT
 10501 GGATATGCAG TTTTCCTAGC ACCATTTAGT GAAGAGGTTG CCTTTTTCT
 10551 ATTAGTGTGT CTTGGCACCT TTGTCAAAAG TCAGTTAGCT GCTATATTCC
 10601 TCCATTTGTG TTGTTATAGA GGAACACATG AGACTAGCAA ATTTATATAT
 10651 CAAATAGAAT TATTTGAATG ATAGTTCTGC ATACTGTACA AGAAGCACAG
 10701 CACTGACTTC TGCTTGGCCT CTGGTAAGGT TCTCAAGATG CTTCCACTTG
 10751 TGGTAGAAGG CAAACATGAG CTGGTATATG CAAAGGTCTC ATGACAAGAG
 10801 AGGAAACCAT AAAGAGGGGA TGTGAGGGAG TGCCAGGTTT TGTAAAACAA
 10851 CTAGCTCTTC TGGGAATAA TAGAGTAAAA ATTCGCCTCC CAGGCAGGGG
 10901 ATTAATCTAT TCGTAGGGA TCTGCTTCCA TGACAAAGGC ACATTCTGTT
 10951 AGATTCTACC CCCAATATTG GGGATCAAAT TTAAACATGA AGTGTGGAGG
 11001 GCTCAAATAT CCATACTATG GCAGCAGTAA ATGCATAAAT TTATTTGTG
 11051 GATCTCTATT CTATATAGTA TTGGTGTATG TATCTGTTT CATGCCACTG
 11101 CCATACTGTT TTGGTGATGA TATCTATGCT ATATATGTGT GTGTGTATAT
 11151 ATATATTATA TATATGTATA TATGTGTATA TTATATATAT GTATATATGT
 11201 GTATATTATA TATATAAAT ACTTTAAGTT TTATATATAT ATAAATACT
 11251 TTAAGTTCAA GGGTACATGT GCAGGATGTG CAGGTCAGTT ACATAGGTAT
 11301 ACATGTGCCA TTTTGGTTTG CTGCATGCAT CAACTCATCA TTACATTAGG
 11351 TATTTCTCCT AATGCTATCC CTCCACCAGC CACCCAACCC CCAACAGGCC
 11401 AGGTGTGTGA TGTCCCCGC CCGTGTGTCCA TGTGTTCTCA TTGTTCACTT
 11451 CCTACCTAAA AGTGAGAACA TGCAGTGTTC GATTTTCTAT CCTTGTGATA
 11501 GTTTGCTGAG AATGACTGTT TTCAGCTTCA TCCATGTCCC TCAAAAGGAC
 11551 ATGAACATCAT CCTTATTTAT GGCTGCATAG TATTCCATGG TGTATATGTG
 11601 CTACGTTTTT TTAATCCAGT CTATCACTGT TGGACATTG GGTGGTTCC
 11651 AAGTCTTTGC TATTGTGAAT AGTGCTACAA TAACCATATG TGTGCATGTG
 11701 TCTTTATAGC AACATGATTT ACTATCCTTT GTGTACATAC CCAGTAATGG
 11751 GATAACTGGG TCAAATGGTA TTTCTAGTTC TAGATCCTTG AGGAATCCCC
 11801 ACATGTCTTC CCACAATGGT TGAATAATT TACATTCCCA CCAACAGTGT
 11851 AAAAACGTTT CTATTTCCCC ACATCCTCTC CAGTATCTGT TGTTCCTGA
 11901 CTTTTTAATG ATGGCCATTC TAACTCACAT GAGATGGTAT CTCATTGTGG
 11951 TTTTGTGTTG CATTTCTCTG ATGACCAAGT ATGATGAGCA TTTTTCATG
 12001 TGTCTTTTGG CTGCATAAAT GTCTTCTTTT GACAAGTGTC TGTTTCATATC
 12051 CTTTGCCAC TTTTCAATGG AGTTGTTTGT TTTTTCCTG TAAATTTGTT
 12101 TAAGTTCATT GTAGATTCTG GATATTAGCC CTTTGTGAGA TGGGTAGATT
 12151 GCAAAAATTT TCTCCCATTC TGTAGGTTGC CTGTTACCC TGATGGTAGT
 12201 TTCTTTTGTGCT GTGCAGGAGC TCTTTAGCTT AATTAGATCC CATTGTGCAA
 12251 TTTGGGCTTT TGTGGCCATT GCTTTTGGTG TTTTAGTCAT GAAACCTTG
 12301 CCCAGGCCCTA AGTCCTCAGT GGTATAGCCT AGGTTTTCTT CTAGGATTTT
 12351 TATGGTTTCA GGTCTAACAT TTAAGTCTTT AATCCATCTT AAATTAATTT
 12401 TTGTATAAGA TGTAAGAAGG GATCCGTTTC AACTTTCTAC ATATGGCTAG
 12451 CGTGTTTTCC CAACACCATT TATTAAATAG GGAATCCTTT CTCCATTTCT
 12501 TGATTTTGTG ATATTTGTCA AACATCACAT GGTTAGAGAT GTGTAGTGT
 12551 ATCACTGAGG CCTCTTTTCT GACTCCATTG ATCTATATAT CTGTTTTGAT

FIGURE 3, page 4 of 11

12601 ACCAATACCA TGTTGTTTTC GTTACTGCAA CCTTGAATG CAATTTGACA
12651 TTCAGGACCA TGATGCCTCC AGTCTCTTTT TTTTCTCTA AATAATTTT
12701 TTGTCAATGT AAGCTCATT TCGCTCTTT CTGATCCATA AAGTATTTT
12751 TTCCCATTTT GTGGAGAACG CCGCNNNNNN NNNNNNNNN NNNNNNNNN
12801 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNGGCACA CCTCGTGCGC
12851 ATATATATAT ATATATATAT ATATACCTCT ATATATATAT ACATACATAC
12901 ATACATACAC ACCTCCTTGT CTGGTGTGGG ATCAGGGTAA TGCTAGCCTC
12951 ACAAGATGAT ACTGAAGTGT TTTTGCCTTT TTGACTTTT GATGGTTTGG
13001 AAGAGTGAGA AAAAGTGTTA TTAATTATTC TTTAAATTTT GTTGAATTTT
13051 ATAGTGAAGA CCTTAGCTCA CTGGCTTTT TAATGAGAAC TTTATTACTG
13101 ATTTAAACTT CTTCTTCATT ATTTATTTCT GCCTTGTTTT TATTTCTTCA
13151 TAATCCAGTC CTATTTTATG TGTCCACTAA ATTGTTTATT TTCCTAGAAT
13201 TTTTCCATTT ATTGGCATAT GCATGTCCAT AGAAGCCTTT TATAGTCCTT
13251 TTCATTTCTA GTGTCATTTT TTTCTTTTT TTTAAGAATC CTTAAGATTT
13301 TAGAGATGAA ATGTCACTTT GTTACGCATA CTGGAGTGG GTGACATTAT
13351 TATAGCTCAC TGAAACCCAA ACTCCTGAGT TTAAGCAATC CTTTACCTC
13401 AAAATTCCAA AATTCCTGAG TAGCTGAGAC AGGCATACAC CATCAAGACT
13451 GGCTAATTTA TTTCAAATTT TGTAAGATG GGTTCCTACT AAGCTATTCT
13501 CAATCTTTGG GCTTCAAGTG ATTCTTCAGC CTCTGTCTCT GAAAATGCTG
13551 GGTTTATAGA TATGAGCCTC TATGCCTGAT TTGCTTTGTC TCTTTGTAAT
13601 CTCCCATTTT ATTTGTGTCT TTTCTGGTTT GTTTCATTTT GTTATGTTTT
13651 CAGTTACCTT GCTAAAGCTT TGTCGATTTT ATCTCTTCAA ACAACTAACT
13701 CAATATTTTG CTGATTTTCC ATATAGTATT TTATTTCTAT TTCATTTATT
13751 TCTGCTCTAA TCTTTGTAA ATATCTTGT TTTCTAATAA TTTTGAGTTT
13801 CCTTGTCTT GTTTTCTAAT TCCTTGCGAT GTTATCATAA ATTGTTTATT
13851 TGATATCTTT CTACTTTTTT GATGTGTGTG TTCGTTGTTG TAGACTTTCC
13901 TCTTTATTAT TCTGATTTCT TCCTCAATTC TCTAATATTA TGATTGCATT
13951 ATTTTCCAAG TTTCTTTTGT TTTTATTTT ATAGTTTATG TGATTCTCTGA
14001 ACTTGTCAAA GAGATTATTG TGAATTTGAT GTCGGATATT TAAGCATTTT
14051 CAAAACCTTT GGTGCATTAT TGAAATTTTA TTGTTTATT TTAGAGATGT
14101 CATACTTCCC AGTTTTTTTT TAACAATACT TGCTCTTTAT ATTGATGTCT
14151 ACATATTTAA AAAGATAACC ACCTGATTCA GCTTTTTTAA GTGATATGCA
14201 GTGGTGTAA GTGTGACTG CTTAATATCA GAGCTGAATC ACTGCCCTGA
14251 GGATTCPTTC TGTTCTGAGG AGAGCTTGTA GTTAATAGCA GAACCTAAAT
14301 AGTGCAGTAG AGCTAAATCT CTTCCATGCT GTTGTTTTCC TGTCTGGGGA
14351 AGACTTATCA TGACCATGAA AACATAATGC TGTGCCAGAA CTTAAACCCA
14401 AACCTGTAGT AATTTCTGAG TTGAGGAAGG CTTAAGAAAT AACTGGAATC
14451 TAGTTACTAA CCTGATAGTT GTTCTGAGT CAGAGAAATG CTCTGCATGA
14501 TCACCTGGGA TATTTGTAAA ATCTAACCAG AGATTCTAGC CTTCCCTTGG
14551 GATTGTGCTT CCTGTACTAC TGTAAGTCTG GCTAGGTCTT CATCAGTGAA
14601 TTCCCTGCTG ATAGGACCAC AAAGCATCTG CCAAGATCTG TTTGCCATTT
14651 GCTGTGATTA GTGCTTCTGC TCTTTGCTTC CAATTCAACT CAGGTGGTTC
14701 AGCCCTTCTG ACACCTCTAA TACCTCCTGT GGGATGGAAC ATAGAAGGCT
14751 TCTCACAATG ATTCACACAC TGATATGGAG ATTGAATGTC CAGTTGCAAC
14801 TATTTTCTTC CACCTGTGTA ATTGCAGGTA CAGGGAAGTT TTCTGTGACT
14851 GATGCTATTT TGGTTTGGAG AATGGGGTGA TGTGGCACA TGATCTTTCT
14901 TCTTTCTGGT CATGGATTTT TTAATTTCCA TGAACCCATA AGATTTTCA
14951 CTTTCTTCT GAGCTCTGGT GCTTTCAGAG TGGTATTTT ATATTGCAAT
15001 AGTTGCTAGT TGTACTTTTA AAAGCGATTG ATGCTGGAGG TCTTCTATTC
15051 CACCATCTCG CTGATGTCAG TCCTCAAATA ATAATTTTAT ATTTTAGCAA
15101 ATTATTTTGG TTTTAGGATT TTGTGTCTAC GTGACACAGA CATGAAAAGA
15151 GATGTACTCA TTAGTGAAAC TTTTTCATA CTGTTTGGT TGTGCGCCTT
15201 TTCTAGTATG AATGATTACA TATTTAAGCC ACATGTTTAA TACATAGACT
15251 GTCCTTTAAA GAGACTAGAT AGTTCTGTGT GTCAGCATAT AGGGACAGAA
15301 TATAACTACA CATTAAATAT TTCTCAAGTA TTTATTTTAG AAGTGTAAGT
15351 AACCTTTATT TTAATTTTGG TTATATTATG CCTCTGTAAT GCAGATAAAT
15401 TTTTATCTTC AGGAAATGGA AAATTTTGTC CAGAGTTCAG GGAAGATGG
15451 TATTGTGGTG TTTTCTCTGG GGTCACTGTT TCAAATGTT ACAGAAGAAA
15501 AGGCTAATAT CATGCTTCA GCCCTTGCCC AGATCCACA GAAGGTGAGT
15551 AAAACCTCCA ATCCTGATAA GCAGCTATTC ACATAATGAA ACAGTATGGT
15601 TTTATTTGGG TCTTGAATCT CATTTTCCAC TTAGCATAAC AGGTACAAA
15651 ATTTGCAAAA CATTATAGTA GTGTACATGG GCATAACTGA TCATTTGCCT
15701 ACTGAGTCTT GCTGTTACTG GAAACAACCT TCTTGATTGT CATTTGTTTA

FIGURE 3, page 5 of 11

15751 TAATAAAAATA GATATAATAA ATAAAGCTCT ACCTTATATT TTAGGATTG
15801 AAATCTAAAA GCGTGTGCCA ATGATPCCAA AAAAAAATTC TGACATCTAT
15851 TATTTCAAAG GACCAGAAAA AGGAAACTG ATATAAAAAA AAAAAAGAAGA
15901 ATCAATCTCA AGAATATCTT CTCATATTTG TGTGTATAAA AACTGTATTC
15951 AGGGTAGTTT TGCTTAGAAA TAAAAGCTCA GATTAATGTA GTCTTTCTAA
16001 ATAATTAGAA GTTTCAAAAG TAAAATGTCA ATTACAATTA TAGTATAGTA
16051 ACAATTATTT AAGTAATGTA ATTATTTATG ATACTCCACT AATTTTAACT
16101 TTATTATTAC TGAATTCTA GAATTTCAACA CTTTAGATAG TGCTATATAT
16151 AAACATATCCA AAAGATATTT CATTTTATAT TTAGCTAAAA TACTTCAAAC
16201 TCAATAAAGG CAAGCATACT AATTAGGAAT TTGAAATATT GTAATTTCAA
16251 TTATGAAATT ATCTGTAAAG TAGTTTGAAA CATCTATGCC GTTCTTTGTT
16301 TTCAAATGTA TAAAATTTGT ATAGGTGTCC AACAAAGAAA AATTGTGTAA
16351 AAAAAAGGTA CAATCTCAAA GAAAATTTAT CATTGAACAG TGGAACATAA
16401 GTAATTTTCT AGCTCATTCT TCTTCAATAA AACAATTAAA TATAAGAAGA
16451 AAGAGGCCAG GAAGGAAATA GAGAAGAAAA GACACCCGAT TATCCAAAAG
16501 ACACACATAA TTGAAAGCAA ATTTTATCT GCAGGGAAC GTAAATTTGA
16551 TGGTAGAATG AGATTGGCTC CATGAGTTAA AATGACACAC AGATCAGGTA
16601 CTTATAAAAT TTTTAATTCT TATATAAAAA TAGATTAGCC ACTGCTGAAT
16651 TATTTTTTTA AATATTCAC GGTATTCTCA TTCTCAAATA TTTTAAATTG
16701 GTAATAAAAT AATAATAGCA TACCTAATAG GCAACTGGTA CACATTATTT
16751 TAAAAGATCT TTGTAAAACG TCCTACTATA TCTTTCAGTC TTTACGCGGT
16801 AGCTCTACAC ACCCCTGTCT CAACCATCAC CTGAAGTACA ATGAGTTTAT
16851 AATTTATAC TATATCTACA TCCTTAGAAT GCTAATATCC TGTGGTTTAC
16901 TCTGTGAAAT ACATGTGTTT CTTCCGTAGG TGTATGGAG GTACAAAGGA
16951 AAAAAACCAT CCACATTAGG AGCCAATACT CGNNNNNNNN NNNNNNNNN
17001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNAAAAAA
17051 AAGACCCAAT CCCAAAGAAA ATTTATCATA GAACAATGGA ACATAGGTAA
17101 TTCTCTAGCT CATTCGTCTT CAATAAAACA AATAAATATA AGAAGACAAA
17151 GGTACAGGAG GAAATAGAGA AGAAAAGATA ACCGATTATC CAAAATCACA
17201 CACAAAATTG AAAGCAAATT TTATCTGTGG GGAAGTGTAA ATTTGATGGT
17251 AGAACCAGAA TAGTTCATG ATTTGAAATG ACACAGAGAT CATGTACTTA
17301 TAAAAATATT TATCTTATA AGAAAATTGA GTAGCCAGTG CTGAATTACT
17351 TTTTAAATGAT TCACTGATAT TCTCATACTC AGATATTTTA ATTGATATTA
17401 AAATAATAAT AGTATACTTA ATAGTCAACT GGTACACATT ATTTGAAAGG
17451 ACTTTTGTA AAGTCCTAC TATGTCTTTT ACTGTTTACA CAGTACCTCT
17501 ACATACCCCT GTCTCAACCA ACACCTGAAG TACAATGAGT TTATAATTTA
17551 TAACTATATC TACATCCTTA GAGTGCTAAT ATCCTGTGGT TCAATCTGTG
17601 AAATACATGT GTTCTTCCA TAGGTGTTAT AGAGATACAA TGGAAAAAAA
17651 ACCATCCACA TTAGGAACCA ATACTCGGCT GTATGATGGG ATACCCCGA
17701 ATGATCTTCT TGGTAGGTCT ATGAGAAAGT AAAAATATGA ACTAGACGAG
17751 GAAAAAATGA ATAAATGTTA AACAGCAAGC AAATTCAGCA AAGATCTAAA
17801 ATTATAAAAC TTTATTTTAC TTAATCTTTT GAAGCAGATA TAATTAAAGG
17851 ATTGACTAAA ATTGTATAGA TTCACACTTT CTATTGTAA GGTGAGAGTG
17901 ACAGGAAATT CAGAAGGAAT TAATGCCTAT TTTTCTGGAG ATAGAAATGA
17951 TCTTTAGTAG CAATGCTCCA TGTGCTCACC TTCTAAAGAA AGTGCTGTAC
18001 GCTTCAGTGA GTTATCTCGT AATTCCCATC TGTAAGTTTT AAATAATTTT
18051 AAAAGTTTAG AATAAAATAT CTCACCATTT CTCATCCAAT TTACATACTA
18101 GGTCAATCCA AAACCAAAGC TTTTATCACT CATGGTGGAA TGAATGGGAT
18151 CTATGAAGCT ATTTACCATG GGGTCCCTAT GGTGGGAGTT CCCATATTTG
18201 GTGATCAGCT TGATAACATA GCTCACATGA AGGCCAAAGG AGCAGCTGTA
18251 GAAATAAACT TCAAAACTAT GACAAGCGAA GATTACTGA GGGCTTTGAG
18301 AACAGTCATT ACCGATTCCCT CGTAAGTACT ACTGCTTGTA CAGACTGATC
18351 TAACATTGAC TATGTTATAC ATTATACCAG AAAATGTTAA ATATCATCCT
18401 GGTAGACATG TTGAGGGATT TTAATCCACA ATATTGAGTC ATTCATCACC
18451 TTGTTACTGG AATAGTTGTG GAAATTGTAG TTCATAGAGT GTCAAACCTT
18501 CTTATGGAAT ATATTAGGTT TAAGTTAACA ACTGGCTTAC TAAGCTTTTA
18551 TTCACATCTT AATTTTACCC CATTTTGTTA AGAATATACT CTTTCAGTCT
18601 CTCCACTATA TCTGTTTAACT ACTATGTAAC CAACAATATT CATGTCACAA
18651 CCAGAATCAA TCTTTTACTG AACATGTTCT TGGCTTGCAT AACATATACT
18701 ACGGTTTATC TACGTGTCTT TTATGAAAAC AAAACTACAA CTTTCTAAGT
18751 TCTATGTGTG TTTTCCCTT CCAGTTATAA AGAGAATGCT ATGAGATTAT
18801 CAAGAATCA CCATGATCAA CCTGTAAAGC CCCTAGATCG AGCAGTCTTC
18851 TGGATCGAGT TTGTCATGCG CCACAAAGGA GCCAAGCACC TGGATCAGC

FIGURE 3, page 6 of 11

18901 TGCCCATGAC CTCACCTGGT TCCAGCACTA CTCTATAGAT GTGATTGGGT
18951 TCCTGCTGAC CTGTGTGGCA ACTGCTATAT TCTTGTTCAC AAAATGTTTT
19001 TTATTTTCCT GTCAAAATTT TAATAAACT AGAAAGATAG AAAAGAGGGA
19051 ATAGATCTTT CCAAAATCAA GAAAGACCTG ATGGGGTAAT CCTGTTAATT
19101 CCAGCCACAT AGAATTTGGT GAAAACCTTG CTATTTTCAT ATTATCTATT
19151 CTGTTATTTT ATCTTAGCTA TATAGCCTAG AATCCACGA TCATGAGGTT
19201 GTGAGTATAT CTCATTCTTT CGTTGTATTT TCCTAGGTGT CTTTACTCTC
19251 TTCTCTCACT TTGTGACACA AGGACATGAA TACATCTAAA TTTTCCTATT
19301 TCTGATATGA CTGTTTTGAT GATGTCATTA CTTCTATAAC CTTAAGTGAT
19351 AGGGTGACAT GCAATATGAT TATTCCTGGT GTGCGCCCAA ACACATGGAT
19401 ATAAAGAGGT AAAAACTTA AAATTCACAA AATTCAGTAA ACCACACAAA
19451 TCAGGTAAGT GTTCTATGAG ATTAGCTGGC TATGAGAAAC ATAATGATGT
19501 TTCTTTTCA ATTTAAATAA GCCCTTCTAC ATAGCCAGCA TCAGTGATCT
19551 CAGAAAATAA ATTGCTAATA ATGATGACAT GGCATTATGC TTAGAAAAGT
19601 TTGCTGTATT TCCATAGACC TCATCTAGAT GTCATGGCCT ACATTTCTGC
19651 CATCACTCAA CCAATACTTT TTTCTGTTTT CTTGATGATA AAAAGACCTT
19701 TCTCATGATT GCCATCAAAT AACAAAAGAA ACTATTTTTT TTCTCACATA
19751 GAGAACATGT TCTCTATGAG TTCAAGGTGA ACAGATATTT TTGGGATTAG
19801 TAACTATTTG AAATATGTGG TGATAATTAC TGAGTTTATA AAATTTATTT
19851 GATAGTACAC TTAAGAAGA TTTATATGTT TATTCTTTAA AAATGATGAA
19901 TACTCATAAT TCTTATCTCT ATAATCAAAA GTATAATTTA CTGTAGAAAA
19951 ATAAAGAGAT GCTTGTCTG AAAGTAAGAT CAGTGAAC TG CTTTTCAGTC
20001 TCAATCTTG AGAATTGTAA ATTCATCAA TAATTGCTTA CATAGTAAAA
20051 ATTTAAGGTA TTAGAAAACC TGCATAACAA ATAGTATTAT ATATTAAATA
20101 TTTTGATATG TAAAGCTCTA CACAAAGCTA AATATAGTGT AATAATGTTT
20151 AACTAGTAA GCAATATGT TAATCTTCTC ATTTTITTTAC TGTATATAA
20201 TCTTAGTGAT ATGCCATTA ATAGTTTAA ATAAATAAAT TGGCTTATCT
20251 GGCTTTTGA AAATTTTGAA ATTCTTACAG ATGTTGATTA GGTATATCTA
20301 CAAATTAATT TCAATTTTAA AATGATGATA TAAAAATAA TATAAGTATT
20351 TTTCTTGTGT ATGTATACAA TAAATATAAA TAAAATTGTT TACTGTTTTG
20401 AAAGTTTCTT AAGTTTTTAC ACTGATATGT TTTTGTACTT TTACAATATT
20451 ATTATAATCT AGGAAAAGCT GATTATATCT GTTTTAAGCC TCATCTTTTC
20501 TCTGTAATTA AACACAGTAA TTTATTAACA TGCTGTGACA GGTGGGAAGC
20551 CATTTCTGGA GTTGAGCCTG CTGACACTCT GGAGCTTTTT AGGTTGGACG
20601 TTCATTGTAT GTGGGACTCT CTGCCTCTCG ATAGCTGTG CTCATAAGAC
20651 TCTCCTTCAT CAATCTGGCA TTGAATTTTG CGATCAGTTG CAATCAGAAT
20701 CCAATTGGCC TTGCCGTTTT AGTATGTTCT ATCTTAACCA GCAATTTCTA
20751 ACCAGGAGCC TGCCAGGTT TGTTCTGTCT TCCCTGTAAG AAGCTCCAG
20801 CATAAATATT CTAAATTTTA CACTACTAAT CTATTAACCA ACCTTTGGAC
20851 CATGTTCACT TTAGGTTGAG CATAGTGTGA TGAGATGCAA ATTAATTTAC
20901 AATCCTATAG GTGTGTGTTA TAAATTTTAA AGTGTATAAA TTAAATAACA
20951 CATTCTAAGT ATCCAACAAA GGTCAAAAAA ATGATATAAA GTCACCAAAC (SEQ ID NO:3)

FEATURES:

Start: 197
Exon: 197-911
Intron: 912-6498
Exon: 6499-6647
Intron: 6648-15412
Exon: 15413-15544
Intron: 15545-16929
Exon: 16930-16940
Intron: 16941-17632
Exon: 17633-17712
Intron: 17713-18101
Exon: 18102-18321
Intron: 18322-18774
Exon: 18775-19051
Stop: 19052

CHROMOSOME MAP POSITION:

Chromosome 4

ALLELIC VARIANTS (SNPs) :

DNA			
Position	Major	Minor	Domain
1735	A	G	Intron
1922	A	G	Intron
2361	C	T	Intron
7371	G	C	Intron
9558	G	A	Intron
10579	T	G A	Intron
10625	C	T	Intron
11147	A	G	Intron
15131	C	T G	Intron
15221	A	G T	Intron
15778	T	C	Intron
15895	-	A	Intron
19786	-	T	Beyond ORF (3')
20157	G	A	Beyond ORF (3')
20246	T	C	Beyond ORF (3')
20681	C	A	Beyond ORF (3')
20819	T	C	Beyond ORF (3')

Context:

DNA

Position

1735	TAC TTTACTGATATACACCTATTTGACTTATTTTATTTCTTTGCTTTACTGATAAAAAG TTGTTTTGCTTTTGCAATTTTCATATAGTTGTGATCAGAGCTGGTCAATGCAAGACATGTT TTTATCCAAATATGTTTGAGAATTATGTAGAAACATGAAAAAGGTACAATTATATCCGA CACTAAAATATTGTTTAATGTATTCCAACGAATTCTTATGCATAGACTGTTTCACAGAAC TAATATTCAGAGGATCCAGTTCAAATGTCCTTAGCCTTAGACATGATTTGAATTTACAT [A, G] TATTGATTTGCTTTAAATAATTTTCCATT CAGTAAGCTGTGCCTAGCTGCAGATAGCCTA CCAGGCTTTATGGATCTAGGTAAACAATACAAATCTCTTGGCCTCAAGTCTACATTCAGA TATTAATTTAAAGGGGTACAGCTATATAGAGGTCACTGGCAAATTTTGGTAAAATAGGAT TATAGTAAAGCCCCCTGACAAGATTGAAATTTAAATAAAACAAAAGTGTATCAAAGG GGTGAAAGAGCATTTTCCAATAAACAAAAGTGGGTTCTGGCCATGCATT CAGAAATTC
1922	ATATTGTTTAATGTATTCCAACGAATTCTTATGCATAGACTGTTTCACAGAACTAATATT CAGAGGATCCAGTTCAAATGTCCTTAGCCTTAGACATGATTTGAATTTACATATATTGA TTTGCTTTAAATAATTTTCCATT CAGTAAGCTGTGCCTAGCTGCAGATAGCCTACCAGGC TTTATGGATCTAGGTAAACAATACAAATCTCTTGGCCTCAAGTCTACATTCAGATATTAA TTTAAAGGGGTACAGCTATATAGAGGTCACTGGCAAATTTTGGTAAAATAGGATTATAGT [A, G] AAAGCCCCCTGACAAGATTGAAATTTAAATAAAACAAAAGTGTATCAAAGGGGTGAAA GAGCATTTTCCAATAAACAAAAGTGGGTTCTGGCCATGCATT CAGAAATTCCTCAACAAT TCTTTAAAAATCATGGAGCAGCTTGATATATAAGAAATTCATTTAATAACTATATTTATT ATGTAGCTCCAACCTACTAAATTATTGATTATTATATATTTTATAGAATTATCTATTGTG AGTCTAAATCAAGAGTATATATTCAAACAATATAGGAAAAGGGATATCAGTCAATTTCA
2361	CAGCTTGATATATAAGAAATTCATTTAATAACTATATTTATTATGTAGCTCCAACCTTACT AAATTATTGATTATTATATATTTTATAGAATTATCTATTGTGAGTCTAAATCAAGAGTAT ATATTCAAACAACCTATAGGAAAAGGGATATCAGTCAATTTCAATTCAGGATTTATTTCC ATAAGTGCTTACGCACAGGTGTATTTCAATTTATTATACATTGCTTTATTGTCCTTCACA AAAATTGCAATTTACAAATTAAAGGTTTTTGAAAACCTGAATCAAGCTAATCAATTTGG [C, T] GTAATATTTCCAACAACAAGTGTGTACTTTTACTCTATCACATATTTGGCATTATCATG CTTTTCAAATTTTTCATTGTTATATCTGTTACGGTGATCTGGGATCAGTGTTCCTTGAT GGTTACACGTTTATTAGCTTGGGGGCACCTTGATGTGTTACAATATAAGACAGCAAACCTT AATTATAAATGTTGTGCATGTACTAACTGCTCCGCTGATTCGTTTCCCCTCCCCTTCT TCTTAGGCCTCCCTATTCCTGAGACACAGTAATATAACATACAATGACTTCTAAATGTT

FIGURE 3, page 8 of 11

7371 AAATAAGTGATGAAGGAAGACAAAGAATAAAAAGAAAATTTCAATAAATGGAAAGTTTAAG
TGTTTAATGATAGTGATGACTTTTACTCAAATAAGTGCTTAGAAGTCATCTGTTTGTGA
TTTATATGATGAATTCTGTGTTGTGACTATCCACTTTGAGCTCGTGAGAATGTAGGTGA
GGTTTAATAAAAAGCCATTTGAGAAAAACAAGGTTTCAACCTCTGTGGACAGAAATCTAAA
TATCGATAGTTATCAGGACAAAGTAGAGCTCATAGAAATAATTTGTCAGCCTGCAGGTTT
[G, C]
TTTTGGAGTGAAAATAAAATTGTATACTATATTCCTAAATCATCAGAGGAAAAATTTAT
AGTTCAAGGAATGTTGAAAGAAACAATATTGAGAAGTAAAAGTGAGTAATAGTTGTTATA
GTTTTTTAATAGTTTTTGTAAAGTATGCTTGAGTTCACTGTCCAAAAGTGGCTATTAGCT
CTAGCCTTGACCTGACAAGGTTCTAGGATATTTAGTCATGGATGTTTATAATCTACCTCT
TACGGGATACTTTTTATTCTGATGAACAGCCTAATGCCTAAGTGTGCAATCTATACCAAG

9558 TCCACATATAAGGGATACCATGCTATCTCTGCCTGGATTATTTCAAGTTAACATTATGCCC
TGGAGGTTCAATCATGTTTCTACAAATGACAGGATTTCAATCTTTTTTTTCCAATATATA
TTTAATGAAATGGATATATATAAACATTGGAAAATGTATATATATATATATCTCCAGT
GGAATGCTATTGAGCTATAAAAAAGTTAATATATAATAGAAATAAAGCTTATATATATCT
AATGGAATGGATATATATATAATGGAATAGAAATATATATCTATACATATAAACACAC
[G, A]
CAATATACATATCCATTTTCAATGTCATATATATATATATAGAGAGAGAGAGAGAGATAT
TTTCAAATGTTGTATATATATCCAATGGAATGGACATATATATATGTATATTTTTTCCA
TATTTTCTTTATGTATTTCTTCATTAATGGATGTTTAGGTTGATTCATCCCTTGGGTATA
TGAATAATGTTGATGTAACATAGAAGGACAGATATCTCTAGACTTCTTAGTTTATTTA
AATATACACCCAGTAATGGAAATGCTGTATAATATGGTAGTTCTATTTTCATTTTTTGAG

10579 CAATGCTTTGATATAACCCATTTATCTACTTTTCTTTTGTGGCTGTGCTTCTGAAGTC
CTATCCAAAAAATCCTTGCTAGACCAATGTCACAAATCATTCCTACAGTTTCTTTC
TAGTAGTTGTATAATGTTTGGCCTTATATTTAACTTTGTAATTCATTTTTACTTACTTTG
TATATGGTGAGGGATAGAGGTCTAGTTTCATTTTCTGCATGTGGATATGCAGTTTCTTA
GCACCATTTAGTGAAGAGGTTGCCTTTTTTCTATTATGTGTTCTTGGCACCTTTGTCAA
[T, G, A]
GTCAGTTAGCTGCTATATTCCTCCATTTGTGTTGTTATAGAGGAACACATGAGACTAGCA
AATTTATATATCAAATAGAAATATTTGAATGATAGTTCTGCATAGTACAGAAGCACA
GCACTGACTTCTGCTTGGCCTCTGGTAAGGTTCTCAAGATGCTTCCACTTGTGGTAGAAG
GCAAAACATGAGCTGGTATATGCAAAGGTCTCATGACAAGAGAGGAACCATAAAGAGGGG
ATGTGAGGGAGTGCCAGGTTTTGTAAACAACTAGCTCTTCTGGGAACATAAGAGTAAA

10625 GTGCTTCTGAAGTCTATCCAAAAAATCCTTGCTAGACCAATGTCACAAATCATTCCT
CCTACAGTTTCTTCTAGTAGTTGTATAATGTTTGGCCTTATATTTAACTTTGTAATTCAT
TTTTACTTACTTTGTATATGGTGAGGGATAGAGGTCTAGTTTCATTTTCTGCATGTGGAT
ATGCAGTTTCTTAGCACCATTAGTGAAGAGGTTGCCTTTTTTCTATTATGTGTTCTTG
GCACCTTTGTCAAAGTCAGTTAGCTGCTATATTCCTCCATTTGTGTTGTTATAGAGGAA
[C, T]
ACATGAGACTAGCAAATTTATATATCAAATAGAATTATTTGAATGATAGTTCTGCATAGT
GTACAAGAAGCACAGCACTGACTTCTGCTTGGCCTCTGGTAAGGTTCTCAAGATGCTTCC
ACTTGTGGTAGAAGGCAAACATGAGCTGGTATATGCAAAGGTCTCATGACAAGAGAGGAA
ACCATAAAGAGGGGATGTGAGGGAGTGCCAGGTTTTGTAAACAACCTAGCTCTTCTGGGA
ACTAATAGAGTAAAAATTCGCCTCCAGGCAGGGGATTAATCTATTTCATGAGGGATCTGC

11147 ACAACTAGCTCTTCTGGGAACATAAGAGTAAAAATTCGCCTCCAGGCAGGGGATTAAT
CTATTCATGAGGGATCTGCTTCCATGACAAAGGCACATTCTGTTAGATTCTACCCCAAT
ATTGGGGATCAAATTTTAACATGAAGTGTGGAGGGCTCAAATATCCATACTATGGCAGCA
GTAAATGCATAAAATTTATTTTGTGGATCTCTATTCTATATAGTATTGGTGTATGTATCTG
TTTTCATGCCACTGCCATACTGTTTTTGGTGTATGATATCTATGCTATATATGTGTGTGT
[A, G]
TATATATATTATATATATGTATATATGTGTATATTATATATATGTATATATGTGTATATT
ATATATATATAATACTTTAAGTTTATATATATATAAAATACTTTAAGTTCAAGGGTACA
TGTGCAGGATGTGCAGGTGAGTTACATAGGTATACATGTGCCATTTTGGTTTGTGCTGCATG
CATCAACTCATCATTACATTAGGTATTTCTCCTAATGCTATCCCTCCACCAGCCACCCAA
CCCCAACAGGCCAGGTGTGTGATGTTCCCCGCCCTGTGTCCATGTGTTCTCATTTGTCA

15131 CAGGGAAGTTTTCTGTGACTGATGCTATTTTGGTTTGGAGAATGGGGTGTGTGGCACAA
TGATCTTCTTCTTCTGGTCATGGATTTTTTAATTTCCATGAACCCATAAGATTTTTTCA
CTTTTCTTCTGAGCTCTGGTGCTTTCAGAGTGGTATTTTTATATTGAATAGTTGCTAGT

FIGURE 3, page 9 of 11

TGTACTTTTAAAAGCGATTGATGCTGGAGGCTTCTATTCCACCATCTCGCTGATGTCAG
TCCTCAAATAAATTTTATATTTTAGCAAATTATTTGGTTTATAGGATTTTGTGTCTAC
[C, T, G]
TGACACAGACATGAAAAGAGATGTACTCATTACTGAACTTTTTGCATACTGTTTTGGTT
GTGCGCCTTTTCTAGTATGAATGATTACATATTTAAGCCACATGTTTTATACATAGACTG
TCCTTTAAAGAGACTAGATAGTTCTGTGTGTCAGCATATAGGGACAGAATATAACTACAC
ATTAATAATTTCTCAAGTATTTATTTTAGAAGTGTAACTTTATTTTAATTTTGT
TATATTATGCCTCTGTAATGCAGATAAATTTTATCTTCAGGAAATGGAATTTTGTCC

15221 TTAATTTCCATGAACCCATAAGATTTTCACTTTTCTTCTGAGCTCTGGTGCTTTCAGAG
TGGTATTTTTATATTCGAATAGTTGCTAGTTGTACTTTTAAAAGCGATTGATGCTGGAGG
TCTTCTATTCCACCATCTCGCTGATGTCAGTCTCAAATAAATAATTTTATATTTTAGCAA
ATTATTTTGGTTTATAGGATTTTGTGTCTACGTGACACAGACATGAAAAGAGATGTACTCA
TTACTGAACTTTTTGCATACTGTTTTGGTTGTGCGCCTTTTCTAGTATGAATGATTACA
[A, G, T]
ATTTAAGCCACATGTTTTATACATAGACTGTCCTTTAAAGAGACTAGATAGTTCTGTGTG
TCAGCATATAGGGACAGAATATAACTACACATTAATAATTTCTCAAGTATTTATTTTAGA
AGTGTAACTAACCTTTATTTTAATTTTGTATATTATGCCTCTGTAATGCAGATAAATTT
TTTATCTTCAGGAAATGGAATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGT
TTTCTCTGGGTCACCTGTTTCAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAG

15778 GTTTCAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCAGATCCC
ACAGAAGGTCAGTAAAACCTCCAATCCTGATAAGCAGCTATTCACATAATGAAACAGTAT
GGTTTTATTTGGGTCTTGAATCTCATTTCCTACTAGCATAACAGGTACCAAATTTGCA
AAACATTATAGTAGTGACATGGGCATACTGATCATTTGCCTACTGAGTCTTGCTGTTA
CTGGAACAACCTTCTTGATTGTCATTTGTTTATAATAAAATAGATATAATAATAAAGC
[T, C]
CTACCTTATATTTTAGGATTTGAAATCTAAAAGCGTGTGCCAATGATTCCAAAAAAAAT
TCTGACATCTATTATTTCAAAGGACCAGAAAAAGGAAAACCTGATATAAAAAAAAAGAA
GAATCAATCTCAAGAATATCTTCTCATATTTGTGTGTATAAAAACTGTATTCAAGGTAGT
TTTGCTTAGAAATAAAAGCTCAGATTAATGTAGTCTTTCTAAATAATTAGAAGTTTCAA
AGTAAATGTCAATTACAATTATAGTATAGTAACAATTATTTAAGTAATGTAATTATTTA

15895 TATGGTTTTATTTGGGTCTTGAATCTCATTTCCTACTAGCATAACAGGTACCAAATTT
GCAAAACATTATAGTAGTGACATGGGCATACTGATCATTTGCCTACTGAGTCTTGCTG
TTACTGGAAACAACCTTTCTTGATGTGCTTTGTTTATAATAAAATAGATATAATAATAA
AGCTCTACCTTATATTTTAGGATTTGAAATCTAAAAGCGTGTGCCAATGATTCCAAAAA
AAATTCGACATCTATTATTTCAAAGGACCAGAAAAAGGAAAACCTGATATAAAAAAAA
[-, A]
GAAGAATCAATCTCAAGAATATCTTCTCATATTTGTGTGTATAAAAACTGTATTCAAGGT
AGTTTTGCTTAGAAATAAAAGCTCAGATTAATGTAGTCTTTCTAAATAATTAGAAGTTTC
AAAAGTAAATGTCAATTACAATTATAGTATAGTAACAATTATTTAAGTAATGTAATTAT
TTATGATACTCCACTAATTTTAACTTTATTATTACTGTAATTCTAGAATTTCACTTTTA
GATAGTCTATATATAAACTATCCAAAAGATATTTCAATTTTATATTTAGCTAAAATACTT

19786 GAAACATAATGATGTTTCTTTTTCAATTTAAATAAGCCCTTCTACATAGCCAGCATCAGT
GATCTCAGAAAATAAATTTGCTAATAATGATGACATGGCATTATGCTTAGAAAAGTTTGT
GTATTTCCATAGACCTCATCTAGATGTCATGGCCTACATTTCTGCCATCACTCAACCAAT
ACTTTTTTCTGTTTTCTTGATGATAAAAAGACCTTTCTCATGATTGCCATCAAATAACAA
AAGAACTATTTTTTTCTCACATAGAGAACATGTCAGTAAGATATTCAAGGTGAACAGA
[-, T]
ATTTTTGGGATTAGTAACTATTTGAAATATGTGGTGATAATTACTGAGTTTATAAAATTT
ATTTGATAGTACACTTAAAGAAGATTTATATGTTTATTTCTTTAAAAATGATGAATACTCA
TAATCTTATCTCTATAATCAAAAGTATAATTTACTGTAGAAAAATAAAGAGATGCTTGT
TCTGAAAGTAAGATCAGTGAAGTCTTTTCACTCTCAATCTTTGAGAATTGTAAATTCAT
CAAATAATTGCTTACATAGTAAAAATTTAAGGTATTAGAAAACCTGCATAACAAATAGTA

20157 ACACTTAAAGAAGATTTATATGTTTATCTTTAAAAATGATGAATACTCATAATCTTAT
CTCTATAATCAAAAGTATAATTTACTGTAGAAAAATAAAGAGATGCTTGTCTGAAAGTA
AGATCAGTGAAGTCTTTTCACTCTCAATCTTTGAGAATTGTAAATTCATCAATAATTG
CTTACATAGTAAAAATTTAAGGTATTAGAAAACCTGCATAACAAATAGTATTATATTA
AATATTTTGATATGTAAAGCTCTACACAAAGCTAATATAGTGTAAATGTTTACACTA
[G, A]

FIGURE 3, page 10 of 11

TAAGCAAATATGTTAATCTTCTCATTTTTTTTACTGTCATATAATCTTAGTGATATGCCTA
TTAATAGTTTTTAAATAAAATAAATTGGCTTATCTGGCTTTTGAAAATTTTGAAATCTTA
CAGATGTTGATTAGGTATATCTACAAATTAATTTCAATTTTAAAATGATGATATAAAAAAT
AAATATAAGTATTTTTCTTGTGTATGTATACAATAAATAAATAAATTTGTTTACTGTT
TTGAAAGTTTCTTAAGTTTTTACACTGATATGTTTTTTGACTTTTACAATATTATTATAA

20246 GAAAAATAAAGAGATGCTTGTCTGAAAGTAAGATCAGTGAAGTCTTTTCAGTCTCAAT
CTTTGAGAATTGTAAATTCATCAAATAATTGCTTACATAGTAAAAATTTAAGGTATTAGA
AAACCTGCATAACAAATAGTATTATATATTAAATATTTTGATATGTAAAGCTCTACACAA
AGCTAAATATAGTGTAATAATGTTTACACTAGTAAGCAAATATGTTAATCTTCTCATTTT
TTTACTGTCATATAATCTTAGTGATATGCCTATTAATAGTTTTAAATAAATAAATTGGCT
[T,C]
ATCTGGCTTTTTGAAAATTTTGAAATCTTACAGATGTTGATTAGGTATATCTACAAATT
AATTTCAATTTTAAAATGATGATATAAAAAATAAATAAGTATTTTTCTTGTGTATGTAT
ACAATAAATATAAAATAAATTTTACTGTTTTGAAAGTTTCTTAAGTTTTTACACTGAT
ATGTTTTTTGACTTTTACAATATTATTATAATCTAGGAAAAGCTGATTATATCTGTTTTA
AGCCTCATCTTTTCTCTGTAATTAAACACAGTAATTTATTAAACATGCTGTGACAGGTGGG

20681 TAAAAATGTTTACTGTTTTGAAAGTTTCTTAAGTTTTTACACTGATATGTTTTTTGACTT
TTACAATATTATTATAATCTAGGAAAAGCTGATTATATCTGTTTTAAGCCTCATCTTTTC
TCTGTAATTAACACACAGTAATTTATTAAACATGCTGTGACAGGTGGGAAGCCATTTCTGGA
GTTGAGCCTGCTGACACTCTGGAGCTTTTTAGGTTGGACGTTCAATTGTATGTGGGACTCT
CTGCCTCTCGATAGCTGTTGCTCATAAGACTCTCCTTCATCAATCTGGCATTGAATTTTG
[C,A]
GATCAGTTGCAATCAGAATCCAATTGGCCTTGCCGTTTTAGTATGTTCTATCTTAACCAG
CAATTTCTAACCAGGAGCCTGCCAGGTTTGTCTGTCTTCCCTGTAAGAAGCTCCCAGC
ATAAATATTCTAAATTTTACACTACTAATCTATTAACCAACCTTTGGACCATGTTCACTT
TAGGTTGAGCATAGTGTGATGAGATGCAAATTAAATTACAATCCTATAGGTGTGTGTTAT
AAATTTTAAAGTGATAAATTAAATAACACATTCTAAGTATCCAACAAAGGTCAAAAAAA

20819 AATTTATTAACATGCTGTGACAGGTGGGAAGCCATTTCTGGAGTTGAGCCTGCTGACACT
CTGGAGCTTTTTAGGTTGGACGTTCAATTGTATGTGGGACTCTCTGCCTCTCGATAGCTGT
TGCTCATAAGACTCTCCTTCATCAATCTGGCATTGAATTTTGCGATCAGTTGCAATCAGA
ATCCAATTGGCCTTGCCGTTTTAGTATGTTCTATCTTAACCAGCAATTTCTAACCAGGAG
CCTGCCAGGTTTTGTTCTGTCTTCCCTGTAAGAAGCTCCCAGCATAAATATTCTAAATTT
[T,C]
ACACTACTAATCTATTAACCAACCTTTGGACCATGTTCACTTTAGGTTGAGCATAGTGTG
ATGAGATGCAAATTAAATTACAATCCTATAGGTGTGTGTTATAAATTTTAAAGTGATATA
ATTAAATAACACATTCTAAGTATCCAACAAAGGTCAAAAAATGATATAAAGTCACCAAA
C